

Factors determining the magnitude of grain-size strengthening in polycrystalline metals

Y. Li,^{a,c} A.J. Bushby^{b,c} and D.J. Dunstan^{a,c*}

^a School of Physics and Astronomy,
^b School of Engineering and Materials Science,
^c Materials Research Institute,
Queen Mary University of London,
London, E1 4NS, England.

Abstract: Grain-size strengthening of polycrystalline metals, the Hall-Petch effect, has been described for the past sixty years as a dependence of the strength of polycrystalline metals on the inverse square-root of grain size, d . The value of the coefficient of the dependence has been the subject of discussion throughout. Here, we find what known factors in the experiments may determine its value, by meta-analysis using maximum-likelihood methods of the literature values of the coefficient in sixty-one datasets. No dependence of the coefficient is found on plastic strain, and a strong dependence is found on the average grain size of each study. Combining these and other factors accounts for the reported values of about 80% of the sixty-one coefficients. The grain-size dependence of the Hall-Petch coefficient is an artefact arising from fitting the data to an incorrect expression. An alternative grain-size effect described by a $\ln d/d$ function is consistent with the theory of dislocations dynamics and generation. The corresponding analysis of the coefficients of fits based on this theory shows that none of the factors investigated are statistically significant, confirming the correctness of this approach.

Keywords:

metals; grain boundary constraint; Hall-Petch effect; dislocations; scaling exponents.

* Corresponding Author

Email address: d.dunstan@qmul.ac.uk
Telephone (+44) 020 7882 3411
Postal address: School of Physics and Astronomy,
Queen Mary University of London,
London, E1 4NS, England.

1. Introduction

The Hall-Petch equation expressing the inverse-square-root dependence of yield or flow strength of polycrystalline metals on the grain size was first proposed in 1951 by Hall [1] and confirmed in 1953 by Petch [2]. Since then, a large body of experimental data has been published in good agreement with this equation. Although it has frequently been challenged, as early as 1958 by Baldwin [3], it is presented in most elementary materials textbooks and university courses and used in industry to predict strength. Most authors consider that it has strong experimental support. Indeed, a conference was held in 2013 to celebrate its sixtieth anniversary [4], with a conference logo incorporating the Hall-Petch equation,

$$\sigma_Y = \sigma_0 + k_{HP}d^{-1/2} \quad (1)$$

In this expression, σ_Y is the yield strength, but the equation has been used as often to describe the flow stress at a given plastic strain, $\sigma(\varepsilon_P)$, and in what follows we will not need to distinguish yield stress and flow stress [5, 6]. The constant σ_0 is the yield or flow stress of single-crystal or bulk large-grain-size polycrystalline material. It is expected to be largely dependent on the history and preparation of the specimen, apart from having, for each metal, a minimum value in ideal specimens due to the Peierls stress. The second term on the right-hand side of Eq.1 describes the dependence of yield or flow stress on grain size d . The notation d_{ISR} will be convenient below for the inverse square-root of grain size, $d^{-1/2}$.

The values of the parameter k_{HP} for different metals have been of great interest throughout this period. Experimentally, values reported in the literature even for the same metal are often very scattered. Many authors have proposed theories consistent with the inverse square-root of d in Eq.1 and capable of explaining the experimental values of k_{HP} . See, e.g., the early review of Li and Chou [7], the comprehensive review of experiments and theories of Cordero *et al.* [8], and references therein, and two modern discussions specifically of the value of the parameter k_{HP} [9, 10]. Such attempts have been frustrated, either by the large scatter in the experimental values of k_{HP} , or by lack of clarity as to the applicability of the predictions of the various theories to specific experimental situations. Thus, some theories predict a non-zero k_{HP} for the yield point. Others predict $k_{HP} = 0$ for the yield point and k_{HP} proportional to plastic strain ε_{pl} for the flow stress. It is debated whether k_{HP} is a derived quantity predictable from more fundamental material parameters, or whether it is a material parameter in its own right to be measured for each metal but not predictable from more basic considerations. Or, like σ_0 , it may be largely dependent on the history and preparation of the samples.

In two previous papers, we have challenged Eq.1 [5, 6]. Following Baldwin [3], we found that the quality of fit of Eq.1 to many datasets was as good with exponents x from $x = 1/4$ to $x = 1$ as it was with the $x = 1/2$ of Eq.1. Using dummy data sets and fitting for the value of the exponent x , we found that the least-square residuals fitting procedure is biased due to the random errors in grain-size estimation, and returns an exponent on average about half the true value. We considered also the probability of the data being where it is under the different hypotheses of Eq.1 and alternative expressions. In the course of that work, we gathered sixty-one datasets from the literature that have been

considered to support Eq.1. In those papers, we concentrated on the raw data, $\sigma(d)$, and whether they supported Eq.1 or were equally consistent with Eq.2 below. Here, we take the opposite approach, setting out to find whether the values of k_{HP} obtained by fitting to $\sigma(d)$ carry any message about the representation of the data by Eq.1 and the physical interpretation of k_{HP} . Note that in this meta-analysis of the sixty-one values of k_{HP} we do not invoke any theories of the Hall-Petch effect. Our purpose here is to find what factors in the experiments influence or determine the observed values of k_{HP} .

We find a clear dependence of k_{HP} on anisotropy, and on stacking-fault energy, and we find no evidence of a dependence of k_{HP} on plastic strain. We find weak dependences on composition (purity) and bulk strength, and on the material parameters determining elastic and plastic anisotropy. Surprisingly, since the dependence of σ on d_{ISR} in Eq.1 is already given explicitly, there is a strong dependence of k_{HP} on d_{ISR} . That is, the coefficient k_{HP} is not a constant with respect to grain size as it should be, but as the function $k_{HP}(d)$ it conceals within itself much of the true functional dependence of σ on d . This and the similar dependence of k_{HP} on the strength, are in fact predictions of the model that we proposed [5,6] in which the Hall-Petch effect is an example of the general size effect, obeying the general size-effect equation [6],

$$\sigma_Y = \sigma_0 + k_{DC} \frac{\ln d}{d} \quad (2)$$

Consequently, we carry out the corresponding meta-analysis of the values of k_{DC} found by fitting the same data to Eq.2. Here, we find no statistically significant dependence of k_{DC} on any of the factors, as expected if this equation does govern the Hall-Petch effect. These outcomes have implications for the theories of Eq.1 for the Hall-Petch effect, which are discussed in Section 4, and provide independent corroboration for the theory underlying Eq.2.

2. Meta-analysis of k values

We apply meta-analysis to the sixty-one values of k_{HP} and of k_{DC} . While commonplace in social sciences and medicine, meta-analysis is relatively unusual in materials science and metallurgy (but see Deville *et al.* [11]). The purpose of meta-analysis is to take multiple studies and by combining their results to obtain a greater statistical significance for a result, or, less often, to obtain a result that the original studies did not consider. Meta-analysis has dangers, which can introduce bias. This is well-documented in the medical literature. See the Appendix for a discussion of their relevance and their mitigation here.

2.1. Data sources and selection.

We assembled a body of data consisting of data-sets that were fitted with Eq.1 by their authors or later authors. Citations, references and search engines led us to more data, as did helpful input from colleagues. We included more recent data when we found it, but our emphasis was on the early data that contributed towards the establishment of Eq.1. All the data-sets that we found, we use; that is, there has been no selection. The sixty-one that we use are distinguished from the unknown number in the literature only by the random accidents of the search processes. There is therefore no risk of selection bias (see

Appendix). We fitted all of the datasets ourselves with Eq.1 and with alternative expressions including Eq.2 [6].

Different authors gave more or less information on specimen characterisation, measurement techniques, and errors, but in any case such information was not used by the original authors to correct in any way the data values fitted with Eq.1 nor the parameter values obtained from the fits. Consequently, it would be inappropriate – it would risk bias – to use any such information here. We work with the raw data.

2.2. Meta-Analysis Factors

Comparing the values of k_{HP} and k_{DC} across many studies requires considering several major factors which are expected from theory to affect values of k_{HP} . Data from both tension experiments and indentation hardness testing are used. We divide the hardnesses by a factor of 2.8 for comparability with the tension data, and attribute the nominal value of plastic strain $\epsilon_{pl} = 0.2$ to these datasets. (While the Tabor factor may be less than 2.8 in some cases, particularly for strain-hardening, that is not expected to affect the meta-analysis significantly). A number of different metals are used. All of the theories of Eq.1 predict that k_{HP} will depend on the elastic moduli E of the metal and on its Burger's vector b . We normalise stresses for different metals by dividing by the Young's modulus to give elastic strains ϵ_{el} . Similarly, the grain sizes are normalised to the size of the crystal unit cell, by dividing the values given by the lattice constant of each metal. (This may be taken as a proxy for normalising to the Burgers vector, which would introduce uncertainties as to the appropriate projections of the vector onto relevant slip planes, etc.) For details of the normalisations see Ref.6 and the Supplementary Information. Following these normalisations, the values of k_{HP} and k_{DC} are dimensionless.

Other known factors in the experiments or theories may contribute to the values of k_{HP} . Some datasets report yield stresses, which ideally would be at a plastic strain of $\epsilon_{pl} = 0$ but may be at the conventional $\epsilon_{pl} = 0.002$ or at a lower or upper yield point, while others report flow stresses at various plastic strains ϵ_{pl} up to 0.3. Different datasets use widely varying ranges of grain size. The metals studied vary in their purity, or number of metallurgically significant elements, from commercial brass and steel to high-purity aluminium. The fitted bulk strength σ_0 may be treated as a factor. In addition, any known physical properties of the various metals can also be considered as factors. Different metals have different elastic and plastic anisotropies and this should affect how polycrystalline specimens behave. We test this by using as factors the elastic anisotropy and the anisotropy of the Schmid factor. We also tested the stacking-fault energy. Finally, the year of publication of the dataset may be relevant, since experimental techniques have changed with time in sensitivity and rigour. So while the year of publication will not be a causal factor, it may be correlated. This brings the total of factors considered here to eight.

In Ref.6, the data were first digitised, normalised as described above, and fitted with Eq.1 and Eq.2. Full information on the datasets, normalisation and fitting are given in [6] and the raw data are given in the Supplementary Information. Here, to test the

effects of these eight factors, the values of k_{HP} and k_{DC} returned by the fits are plotted against each factor in turn, and tested for correlation with each of the eight factors by fitting the data to the function $y = ax + b$. Independence of k of a given factor corresponds to fitted values of a consistent with zero within the statistical error bar. In almost every case we find a statistically significant non-zero a . So we then test the factors acting together.

2.3 Statistical Methods

In a standard analysis of experimental results, data are obtained as a function of experimental parameters in the light of theory. There will be a predicted functional dependence and perhaps quantitative predictions of coefficients, and it is to test and refine these predictions that the experiments are performed. Meta-analysis proceeds differently. We have sets of reported data, here, values of k_{HP} , and of k_{DC} , and potential factor values, constituting a large matrix of numbers. The objective is to establish correlations within this matrix. The most powerful way to do this is factor analysis but for our purposes here it is preferable to use a less powerful but more transparent technique.

We begin by inspecting the properties of the 61 normalised values of k_{HP} . They have a mean of 0.155, but a wide distribution of values from -0.001 to 0.998 (Fig.1), so the mean is near the lower end of the range. The standard deviation about the mean is 0.207. The kurtosis (fourth moment over second moment) is 11 and the skewness is 3, compared with the values 3 ± 0.5 and 0 ± 0.3 expected for a Gaussian distribution of 61 numbers. For $\log_{10}k_{\text{HP}}$, the mean is -1.02 , standard deviation 0.214, kurtosis 7.5 and skewness -1 . The values of k_{DC} have similar properties, with a distribution of values from 0 to 18, mean of 1.49, standard deviation 2.06, kurtosis 24 and skewness 4.3.

The high values of kurtosis indicate that the distributions of the data are far from normal or lognormal. Least-squares fitting methods assume that the residuals r_i – the scatter of the data around the fitted model – are Gaussian-distributed i.i.d. (independently drawn from an identical distribution). When that is not so, as here, least-squares methods discard much of the information in the data, and it is preferable to use other methods which make use of more of the information by using more appropriate probability distribution functions (pdf). We use Maximum Likelihood methods. For accessible introductions to these methods, see e.g. [12, 13].

Fig.1 near here

We proceed by selecting a suitable pdf by testing against dummy datasets. Suitable dummy datasets may be constructed as (ε_i, k_i) where ε_i are random numbers in the range 0 to 1, and k_i are the Hall-Petch parameters ($i = 1$ to 61). These also provide a benchmark for what we should expect when the Hall-Petch parameters have no dependence on, nor correlation with, a factor. Using a least-squares procedure to fit these dummy data to $ax + b$, we expect, within error, $a = 0$ and $b = \bar{k}$ where $\bar{k} = 0.155$ is the mean value of our collection of k_{HP} values. One such dataset gives $a = -0.056 \pm 0.098$, as expected consistent with zero within error, and $b = 0.182 \pm 0.054$ (Fig.1a). The value of

b should be the mean of the values of the data; however, it has an uncertainty because the correlation matrix for a and b has off-diagonal terms of about 0.9. It is better to fit to the datasets $(\varepsilon_i - \bar{\varepsilon}, k_i - \bar{k})$, and restore the average values after fitting. This gives the same result for a , with $b = \bar{k} - a\bar{\varepsilon} \pm 0.027$ and the off-diagonal terms in the correlation matrix are now zero. In this way, we created 500 such dummy datasets. The 500 values of a averaged to -0.002 and the standard deviation of the values of a was 0.055 (Table I). The value of b is no longer interesting, as the fit necessarily goes through the point $(\bar{\varepsilon}, \bar{k})$, so there is only one free fitting parameter, a .

[Table 1 near here. See page 21]

Given the presence of excess kurtosis, the few datapoints at values of k_{HP} up to 1 or values of k_{DC} up to 18 heavily distort the least-squares fits. A common procedure in such cases is to eliminate these datapoints, as outliers, and fit to the remaining points which have a much narrower and approximately Gaussian distribution. It is better not to delete any data, but to exploit all the information in the data using Maximum Likelihood methods, in which any probability distribution function (pdf) may be used as best fits the data. For any pdf $P(r)$ for the residuals r_i , the likelihood of each datum k_i is $P(r_i)$, and the likelihood of the dataset under the model is

$$L = \prod_{i=1}^n P(r_i), \quad \ln L = \sum_{i=1}^n \ln P(r_i) \quad (3)$$

Residuals are functions of the parameters a and b of the model, and for a normal distribution of residuals centred on zero, the standard deviation σ of the Gaussian pdf is the third parameter of the model, with a and b . For the reader familiar with conventional least-squares (LS) but unfamiliar with maximum-likelihood (ML) analysis, it is worth expanding Eq.3 for a Gaussian pdf,

$$P(r_i) = \frac{1}{\sqrt{2\pi}\sigma} e^{-r_i^2/(2\sigma^2)}$$

$$\ln L = -\frac{1}{2\sigma^2} \sum_{i=1}^n r_i^2 + \frac{n}{2} (2 \ln \sigma + \ln 2\pi) \quad (4)$$

Thus maximising $\ln L$ is the same thing as minimising the sum of the squares of the residuals. Least-squares methods are simply maximum-likelihood methods under the assumption (usually false) that the residuals are i.i.d. from a Gaussian distribution. The likelihood L is the probability of the data given the model or hypothesis and it is the key term entering into the Bayes' Theorem expression for the likelihood of the hypothesis given the data.

Values of L can be very large or very small, depending on the width of the pdf. It is more convenient to sum the natural logarithms of $P(r_i)$ to calculate the log-likelihood, $\ln L$. This is then maximised with respect to a , b and σ . Note that the absolute values of L or $\ln L$ have no intrinsic meaning: it is changes of these values between models that matter. Maximising $\ln L$ with the single dummy dataset mentioned above and a Gaussian pdf, we get $a = -0.056$, as above, and $\sigma = 0.29$ which is not an error estimate for a or b but describes the variance (σ^2) of the data with respect to the model. The log-likelihood is

$\ln L = 10$. Error estimates for the fitting parameters can be found by varying each individually while leaving the others as free fitting parameters and looking for a reduction in $\ln L$ of 0.5. Here that gives $a \pm 0.096$ in agreement with the least-squares method, as expected for a Gaussian pdf. The significance of $\Delta \ln L = 0.5$ is that it corresponds to a one-sigma preference in traditional statistics [12, 13].

We calculate $\ln L$ with other pdf's. Given the large kurtosis, we are interested in fat-tailed distributions compatible with a significant proportion of apparent outliers in the dataset. The Lorentzian is suitable, with a width γ . The maximum of $\ln L$ is found to be 56.5 for $a = -0.018 \pm 0.024$, with $\gamma = 0.034$. Thus a is found to be significantly closer to zero, by a factor of about three, than by the least-squares model (Fig.1a). The increase in $\ln L$ is massively significant – with 0.5 in $\ln L$ corresponding to 1σ in a Gaussian analysis, here we have an increase of 46.5, or 93σ . Moreover, we do not have to consider that the datapoints that are a long way from the fitted line are outliers to be discarded – on the contrary, they are accommodated by the Lorentzian pdf on a par with all the other datapoints.

Other fat-tailed pdfs give similar results. We can attribute the apparent outliers to a broad Gaussian pdf and the other data points to a narrow one about the line $y = ax + b$, so that we have a double-Gaussian pdf, referred to below as DG. Then, as well as looking for the values a , b and the σ -values of the two distributions, we look for the fraction f of the broad pdf and the fraction $(1 - f)$ of the narrow pdf that maximizes $\ln L$. Fitting with the two extra parameters, the width of the second Gaussian and the value of the fraction f , we obtain $a = 0.010 \pm 0.025$, $f = 0.142 \pm 0.050$, and the standard deviations of the narrow and broad distributions are 0.067 and 0.80 respectively. That is, about one seventh of the data belong to the broad distribution, and the others form a distribution sharper by a factor of more than two than the single Gaussian pdf gave. The value of $\ln L$ is 61, substantially above the Lorentzian value. Each extra fitting parameter requires an increase in $\ln L$ of at least 1 to offset it (this corresponds to the Akaike information criterion [14]), so this model is only slightly preferable to the Lorentzian. However, it provides a first insight into the number of apparent outliers – the fraction f of the data, or about ten datapoints that appear to belong to a different pdf.

Alternatively, we can consider that each datum has a probability f of being an outlier in this sense, and that the outliers have a flat distribution over a range of k values of width g , hence a flat pdf of magnitude fg^{-1} over the range. Then $(1 - f)$ of a sharp Gaussian distribution is added to make the Gaussian plus flat (GF) pdf. Now, with $g = 1$, maximization of $\ln L$ yields $a = 0.011 \pm 0.025$, $f = 0.158 \pm 0.056$ and $\ln L = 68.5$ (Fig.1a). Especially with one fewer parameter than the double Gaussian, this increase in $\ln L$ is significant – this is the preferable model.

Similar results are obtained by fitting to values of $\log_{10} k_i$ (Fig.1b). The ordinary least-squares method works better here, as there is less skewness in this distribution, but still gives an error in the gradient nearly three times the gradient given by the GF pdf. The least squares $\ln L$ is -39 , rising to -29.5 for the Lorentzian, -26 for the double-Gaussian and -25 for the GF pdf. (Recall that the absolute value of $\ln L$ is not important;

it is lower here because the flat part of the GF pdf is spread more thinly on the y-axis from -3 to 0 instead of 0 to 1 , corresponding to $g = 3$ in Table 1). The main difference is the much higher proportion of outliers attributed to the $\log_{10}k_i$ by the DG and GF pdfs. It probably occurs because the true outliers are much closer to the main distribution, as seen in the DG results by the relative values of σ_1 and σ_2 .

What we have established in this Section is that the Maximum Likelihood methods are about four times more sensitive than the least-squares method for exposing a correlation or lack of correlation between the data, the experimental values of k_{HP} , and the abscissa or factor against which they are plotted. The key benchmark is the standard deviation of the fitted values of the slope, a . Using least-squares fitting, a slope a of 0.096 (on a plot where the abscissa values have been normalized to the range $0-1$) is not significant. Using Maximum Likelihood methods, this criterion is sharpened to about 0.024 . Of these methods, the flat distribution for the outliers with a Gaussian for the bulk of the data is the most probable model. There is a consensus among the methods that about 15% of the data, or ten of them, are outliers. The GF pdf gives the highest log-likelihoods and so it is results with this pdf that are reported below in Tables 2 and 3. The k_i and the $\log_{10}k_i$ give very similar results, except that the outliers are probably over-reported in the $\log_{10}k_i$ analysis.

Similar results are obtained for k_{DC} . This is not surprising, since the kurtosis and skewness are similar, so in what follows we use the same methods. For comparison with other fits below, the equivalent dataset to Fig.1a with a GF fit gives $\ln L = -89$.

3. Fitting to the Experimental Factors

The next step is to plot the data against the various physical factors reported by the original authors and to fit with $ax + b$ as for the dummy datasets of the previous Section.

3.1 Fitting to single factors

In order to include all sixty-one data in every fit and maintain comparability between the $\ln L$ values, the missing values for plastic strain (brass), anisotropy (Ti) and the stacking fault energy (Cr) were allocated mid-range values (in italics in the Supplementary Information) which should not affect the gradients significantly. For comparability, the range of each factor is rescaled to the dimensionless range $0-1$. The resulting gradient a for each factor is given in Table 2a for the fits of k_{HP} to each of the eight factors. The corresponding gradients α for fits to the real values of the factors are also given; where a is negative, we reverse the range of the factor, $(1-0)$ to have a positive α . It should be noted that the slopes α are dimensionless (except for stacking-fault energy and year of publication) because of our initial normalizations of yield or flow stress by division by the Young's modulus to give yield or flow elastic strain, and grain size by division by the lattice constant to give a number per grain. These initial normalizations are essential if different metals are to be compared.

Each fit reported here for Hall-Petch data from Eq.1 was also performed using the parameter values k_{DC} and σ_0 returned by fitting the data to the $\sigma = \sigma_0 + k_{DC} d^{-1/2} \ln d$

function corresponding to the general size effect theory [6]. The results are reported in Table 2b.

Fig.2 near here

The fits of k_{HP} against the normalized plastic strain ϵ_{pl} , are shown in Fig.2a (data from 0.2 to 1 not shown but included in the fit) and the fits to k_{DC} in Fig.2b (data from 2 to 20 not shown but included in the fit). The results with the two maximum-likelihood pdfs shown are very consistent in giving a positive slope marginally outside error at $1\frac{1}{2}\sigma$. At less than 2σ , this might be interpreted as evidence for no effect of strain on the Hall-Petch parameter, and the increase in the k_{HP} log-likelihood to 70 (Table 2a) from the random-variable value of 68.5 (Table 1) offers scant support. The value of α_ϵ for this fit is meaningful, and dimensionless because k and strain are both dimensionless. Similarly, for k_{DC} the lack of change in $\ln L$ from the value of -89.3 for the random abscissa is not significant.

To describe the purity or composition of the metals studied, given the complexities of metallurgy, we adopted the simple scheme of assigning the value $p = 0$ to a pure metal (four or five nines), $p = 1$ for the addition of an alloying element (as in brass) or purities around two nines, and $p = 2$ for anything more complicated, i.e. commercial iron and steel. There is a strong effect, with gradients a significant at over 4σ (Table 2). The jump in the k_{HP} log-likelihood to 76 (Table 2a) is significant. The value of α for this fit is not important because of our arbitrary quantification of purity p . Similarly, for k_{DC} the increase in $\ln L$ to -82 (Table 2b) from the value of -89.3 for the random abscissa is quite significant. Alloy effects are not unexpected, because of the many other ways alloying can increase strength, potentially screening the grain-size effect, or, as here, apparently increasing it by increasing k_{HP} .

[Table 2 near here. See page 22]

For the bulk strength, we used the values of σ_0 obtained by fitting the data to Eq.1 (for k_{HP}) and Eq.2 (for k_{DC}) (see Supplementary Information). Because σ_0 is normalized as elastic strain, the value of α for this fit, α_σ , is dimensionless. For elastic anisotropy, we used a mid-range value for titanium, as the anisotropy factor is less well defined for hexagonal metals than it is for the cubic metals. For the cubic metals, we used the standard definition that is based on the ratio of the two shear moduli, $C = 2c_{44}/(c_{11} - c_{12}) - 1$. This has the advantage, compared with e.g. $C = 2c_{44} - c_{11} + c_{12}$, of being already normalised for different metals. It shows a stronger effect, at almost 5σ , but a fallback in $\ln L$ to 71 (Table 2). Because C is dimensionless, the value of α for this fit, α_C , is dimensionless. For k_{DC} the lack of change in $\ln L$ from the value of -89 for the random abscissa shows that purity is not a significant factor here.

Fig.3 near here

Estimates of stacking-fault energies γ were found in the literature for all the metals except Ti and Cr (see Supplementary Information). Their slope against γ is

negative for both k_{HP} and k_{DC} , so for easier comparability with the other factors, we used the reverse of the stacking-fault energy, $\gamma_{\text{max}} - \gamma$. The slopes α_γ for this factor have units of $\text{m}^{-2} \text{mJ}$. For k_{HP} , the slope α_γ has the largest significance, at 6σ , and for both k_{HP} and k_{DC} the increases in $\ln L$ above the dummy value are the largest, at 15 and 8 respectively.

For the plastic anisotropy, represented by the anisotropy ΔM of the Schmid factor M , we have only the three values of ΔM for BCC, FCC and HCP metals. So this factor may be a proxy for any other ways these classes of metals may differ, [including propensity to develop texture, and any other effects correlated with crystal symmetry](#). The slope α_M for this factor is dimensionless. For k_{HP} the $\ln L$ is only 69 for this factor (Table 2a) and for k_{DC} , -87 , scarcely above the random abscissa values; however, the gradients are non-zero to about 3σ .

For the grain size (Fig.3), we take the mid-points of each dataset on the classic Hall-Petch plots, i.e. $d_{\text{ISR}}^{\text{mean}} = 1/2(d_{\text{ISR}}^{\text{max}} + d_{\text{ISR}}^{\text{min}})$. Because the grain size is normalized to the number of lattice constants per grain, the value of α for this fit, α_d , is dimensionless. This characterization of d is motivated by consideration of what a d^{-1} dependence looks like on a Hall-Petch plot. It is a parabola, and the slope of a straight line fitted to data that follow the parabola would be approximately the gradient of the parabola at the mid-point of the data range. Of course, any dependence of k on any measure of d would contradict Eq.1. Yet this factor has strong effects for k_{HP} , with a very significant slope, at 5σ , and with a jump in $\ln L$ to 74 (Table 2a). On the other hand, for k_{DC} , the grain size gives a $\ln L$ of only -87 , close to the random abscissa value (Table 2b).

Data comes from papers with a wide range of publication dates, from 1919 onwards. The slope a against year is negative; we reverse this for a positive α by plotting the data against the age (2018 minus publication year). The slope α has the units yr^{-1} . Since the year can have no causal influence, it is not surprising that the slopes for k_{HP} and k_{DC} are both small, and not very significant. The $\ln L$ is among the lowest in both cases, and not very significantly above the random abscissa values.

3.2. Combining Factors

Compared with the null hypothesis of Section 2.3.1, i.e. the hypothesis that there is no effect of these factors on the Hall-Petch parameter k so that random ordering would give the same results as ordering the k values by these factors, all eight of these models are supported by the data having gradients more or less significantly different from zero in fits to the data using k_{HP} (Table 2a) and in fits using k_{DC} (Table 2b). We should consider the effect of these factors acting together.

It is worth considering how this would work in the ideal case. Suppose that these eight factors are the factors fully determining k , that they are independent, and that their linear ($ax + b$ with $b = 0$) contributions to k simply sum. Then in the fits to single factors so far discussed, the other seven factors contribute both the intercepts b and the scatter which gives the uncertainties in a in Table 2. If we suitably scale all eight factors and add

their contributions to k , the intercept b should decrease and the uncertainties in a should decrease; a itself should increase. Keeping the abscissa normalized to the range 0 to 1, with all eight factors included, the slope a should be within error equal to the highest values of k not belonging to outliers, i.e. about 0.18 for k_{HP} and about 2 for k_{DC} , and perhaps more if some of the outliers are brought within the main distribution when all factors are considered. The remaining outliers will be the result of other large, rare perturbations due to unidentified factors.

The combined abscissa with all positive coefficients α is $\alpha_Y (2018 - Yr) + \alpha_C C + \alpha_d d_{\text{ISR}} + \alpha_\varepsilon \varepsilon_{\text{pl}} + \alpha_p p + \alpha_\sigma \sigma_0 + \alpha_\gamma (\gamma_{\text{max}} - \gamma)^{-1} + \alpha_M \Delta M$ for k_{HP} ; the grain size term for k_{DC} also becomes $\alpha_d (d_{\text{ISR max}} - d_{\text{ISR}})$. For comparison with the foregoing single-factor fits and plots, we rescale this new abscissa again to the 0-1 range. The results of fitting to the k_{HP} dataset are given in Table 3. The slopes a are much increased, to about 0.15 and 1.75, nearly to the expected values of 0.18 and 2. Most significantly, the $\ln L$ increases sharply, for k_{HP} , to 92 from the values 69 to 76 for the factors taken singly (Table 3a). This is a very much more preferable model than any of the eight factors taken alone. Similarly, in Table 3b, we see that combining the eight factors in the same way for k_{DC} gives a substantial improvement, with $\ln L = -72$ falling well above the range of the values of -89 to -81 for the factors taken singly.

However, we improve $\ln L$ further if we multiply the factors by weightings w_i . Eight weights w_i are actually only seven new parameters, since the sum of the eight is not important – indeed either it or one of the weights has to be specified before the $\ln L$ can be maximized with respect to the eight weights w_i as well as to a, b, f and σ . For k_{HP} , the optimum combined abscissa in Fig.4 uses the weightings of the third column of Table 3a. This gives another large increase in the log-likelihood to 101, a slope a of 0.156, and a very small intercept of $b = 0.01 \pm 0.01$. The errors on the weightings are obtained as above, by fitting with each weight fixed in turn at its optimized value plus or minus an offset, and looking for the value of the offset which reduces $\ln L$ for the fit by 0.5. In contrast, for k_{DC} , the optimized weightings give no significant improvement in $\ln L$ over the equal weightings.

[Table 3 near here. See page 23]

[Fig.4 near here]

Increasing the number of parameters by seven means that we cannot compare the values of $\ln L$ directly. Any extra parameter in fitting will normally give an improved fit, but multiplying parameters unnecessarily is to be discouraged. This is the purpose of the Akaike Information Criterion, $\text{AIC} = 2n - 2\ln L$, and the Bayesian Information Criterion, $\text{BIC} = n \ln m - 2\ln L$, where n is the number of parameters and m is the number of data. With 61 data in our case, $\ln m = 4.1$, so the BIC discourages extra parameters more strongly than the AIC. If adding an extra parameter does not decrease the AIC or BIC by more than 2, that parameter may be considered to be unnecessary. We can now find out which of our eight factors may be unnecessary by deleting each in turn (setting its weight to zero), and then maximizing $\ln L$ against the other parameters). Results are given in

Table 3a for k_{HP} . Deleting the year of publication as a factor leaves $\ln L$ unchanged and improves both the AIC and BIC significantly. Deleting the strain reduces $\ln L$ marginally, but the small changes in the AIC and BIC do not justify keeping this factor. Similarly for the other factors, except the elastic anisotropy and the grain size, and perhaps the stacking-fault energy. What is most dramatic here is that deleting the grain size causes a dramatic worsening of the BIC, an increase of 20. In terms of $\ln L$ alone, the decrease of $\ln L$ by 22 translates into a probability of the model including grain size which is $\sim e^{22} \sim 3 \times 10^9$ higher than the probability of any of the models k_{HP} that do not include it as a factor.

In contrast, the results given in Table 3b for k_{DC} show that none of the parameters are necessary. Neither adding the weights, nor deleting each factor in turn by setting its weight to zero, affects the $\ln L$ significantly. Consequently, the AIC and BIC are uniformly worse (higher) by about 10 and 20 respectively compared with the fewer-parameter equal-weight model. This shows that none of the factors are necessary in models of k_{DC} .

4. Discussion

Several points stand out among the results presented here. The random ordering of the datapoints on the abscissa of Fig.1 gives a slope a which is not significantly different from zero. About one-third of the possible random orderings will give a more than one standard deviation σ from zero, and about one-twentieth will be more than 2σ from zero. In contrast, of the orderings of the data given by the various factors, Table 2a shows that all but one give an a more than σ from zero and all but two are more than 2σ from zero. Clearly, there are meaningful correlations here between the factors and the reported values of k_{HP} . Correlation, however, is not causality. These factors have correlations among themselves, both accidentally and intrinsically. Plastic strain ε_{pl} and bulk strength σ_0 provide an example of an intrinsic (positive) correlation arising through strain-hardening. Year of publication and our purity factor p provide an example of an accidental (negative) correlation, because the metals to which we have given the highest p values are brass and steel, which were the first to be reported (1919 and 1951, respectively), while very pure metals, with the lowest p value, were reported much later.

Given these correlations, which are likely to link any of the factors considered to whatever factors are causally related to k_{HP} and k_{DC} , it is not surprising that most of the gradients a for single factors (Table 2) are significantly different from zero. Nor is it surprising that combining the factors improves the fits to the data (Table 3, columns 2) and that giving the additional degrees of freedom by variable weights can improve the fit further. The difference between seven and eight non-causal accidental correlations should not be large. It is not surprising that deleting a single factor generally has little effect on the $\ln L$, AIC and BIC. The weights of the other factors can adjust to largely compensate for the deleted factor. In Table 3a, the only factor for which this is not true is the grain size. Deleting any of the other seven parameters leaves $\ln L$, AIC and BIC values within the range of the equal-weighting (column 2) and optimized weighting (column 3) values. But deleting the grain size has a dramatic effect, leaving $\ln L$, AIC and BIC all very

significantly worse. The conclusion is that the grain size is the only necessary factor, i.e. the only factor with a direct causal effect on the value of k_{HP} . This is complete contrast to Table 3b, where any of the parameters including grain size can be deleted with little effect on the $\ln L$, AIC and BIC.

The necessary conclusion from the comparison of Tables 3a and 3b is that the grain size is an independent parameter entering into k_{HP} but not into k_{DC} . It will also enter into both through its accidental correlations with other factors determining both k_{HP} and k_{DC} , but its independent contribution to k_{HP} invalidates Eq.1, the classic Hall-Petch equation.

That leaves the question of what other factors do give rise to the variations of k_{HP} and k_{DC} that are observed. Many theories of Eq.1, and the theory of Eq.2, have k_{HP} and k_{DC} as constants, when normalised as we do throughout this paper, and the theoretical constant value of k_{DC} is close to the mode of the GF distribution [6]. So, finally, we consider why the reported values of both k_{HP} and k_{DC} are so widely scattered. There are two separate questions, why about 85% of the data are distributed from zero to 0.2 (k_{HP}) and from zero to 2 (k_{DC}), and why 15% are distributed up to 1 (k_{HP}) and up to 18 (k_{DC}). To address the first question, note that in very few of the 61 datasets is the strength σ varied by as much as a factor of two between the largest and smallest grain sizes, and often the range is much less than that. In other words, the grain-size strengthening is generally much less than the bulk strength σ_0 in Eq.1 and Eq.2. The datasets are generated by preparing specimens with widely different grain sizes resulting from widely different annealing conditions. It is difficult to imagine that the bulk strength is unaffected by the annealing. Depending on the metallurgical complexity of the specimens, annealing can cause softening or hardening independent of the grain growth. It would require only a small change in bulk strength, correlated with the grain growth, to change k_{HP} dramatically. In Ref.6, Fig.1 shows datasets for nickel and aluminium where the bulk flow or yield elastic stress exceeds 10^{-3} while the normalized grain size is around 10^5 . In the case of the aluminium, an increase of less than 10% in σ_0 under the conditions that take the normalized grain size from 4×10^4 to 7×10^5 would be sufficient to account for the fitted values of k_{HP} being close to zero and even negative as observed. In the case of the nickel, similarly, a still smaller softening under the conditions that take the normalized grain size from 6×10^4 to 2×10^5 would be sufficient to account for the fitted values of k_{HP} being at the high end of the range, near 0.2 as observed.

To address the second question, from Fig.4, we note that the 15% of the data associated with the flat (F) of the GF pdf, which we may call outliers, are almost exclusively from commercial brass and steel, and from the early years of investigation of the Hall-Petch effect. These are the most complicated materials studied, metallurgically. The grain-size measurements were done on them by the classic but difficult techniques of revealing grain boundaries by polishing and etching and observing them under an optical microscope. We can only speculate that there were sometimes systematic errors (not always, for some of the brass and steel data are not outliers), i.e. failures to observe features of the microstructure smaller than the measured grain size but correlated with it.

The variation of reported (normalised) values of the Hall-Petch coefficient k_{HP} that fall in the range 0 to 0.2 is largely accounted for by the various factors reported by the original authors – specifically, by the purity, the bulk strength and the grain size, and also by the material parameters, the elastic and plastic anisotropy and the stacking-fault energy. That the plastic strain is not required as a factor is significant for many theories of the Hall-Petch effect. But none of these factors except grain size enters in causally: there are only accidental correlations. That grain size itself is one of the factors that matters most, unambiguously demonstrates that fitting the data to Eq.1 is incorrect.

Some comment can be made on the consequences of these conclusions for the various theories of the Hall-Petch effect. It is now clear that the difficulties previously encountered in explaining the values of the Hall-Petch coefficient k_{HP} of Eq.1 are due to trying to explain an incorrect equation. The theories that invoke plastic strain will need to be reconsidered, perhaps in terms of the cluster of correlated factors more-or-less related to strength. The theories that do not invoke the factors found here, such as pile-up [1, 2, 15], are very restricted in their application, since they would account for an intercept on the ordinate of Fig.4, which here is zero within error. Grain size is the dominant factor for k_{HP} , which is a clear demonstration that Eq.1 is incorrect. That the critical thickness theory constant k_{DC} does not depend on any of the factors considered here, especially not grain size, strongly supports the theory of Eq.2, the critical thickness theory that invokes dislocation curvature or source size constraint. [This theory is explained in detail elsewhere \[6, 27, 28 and references therein\]. In summary, it is based on the observation that any dislocation configuration under stress has curvatures proportional to the stress, so that if the configuration is rescaled to another size, the curvatures and the stresses vary with the inverse of the size. This can be applied to situations as diverse as the generation of misfit dislocations at a misfitting interface, and Frank-Read or spiral source operation in the strain gradient of a wire under torsion.](#)

[The unexplained variations in the reported values of both \$k_{HP}\$ and \$k_{DC}\$ cannot be attributed to any known physical factors, but only to systematic errors in the experimental design, such as dependence of the bulk strength on annealing conditions.](#)

5. Conclusions

Meta-analysis of a large amount of data purported to support the Hall-Petch equation, Eq.1, clearly shows that the Hall-Petch coefficient k_{HP} is not a material constant, nor a function of experimental parameters such as plastic strain, but it is strongly dependent on grain size. This demonstrates that the long-held HP relationship is not the correct description of grain-size strengthening in metals. An alternative relationship, Eq.2, is founded on the relationship between stress and curvature of dislocations and the constraints that size puts on curvature in dislocation configurations such as dislocation sources [27, 28]. This does give a sound description of the grain size strengthening with no interdependence of the terms in the equation, σ_0 , k_{DC} and d – each of which can be justified independently.

Acknowledgements

Yuan Li is grateful to the Chinese Scholarship Council for his research studentship. This research did not receive any specific grant from funding agencies in the public, commercial, and not-for-profit sectors.

Declaration of Interest - None

References

- [1] E.O. Hall, The deformation and ageing of mild steel: III Discussion of results, *Proc. Phys. Soc. B* 64 (1951) 747–753.
- [2] N.J. Petch, The cleavage strength of polycrystals, *J. Iron Steel Institute* 174 (1953) 25–28.
- [3] W.M. Baldwin, Yield strength of metals as a function of grain size, *Acta Metall.* 6 (1958) 139–141.
- [4] International symposium on Strength of Fine Grained Materials - 60 years of Hall-Petch, Tokyo, 2013. URL: <http://www.u-gakugei.ac.jp/~tkosaka/sfgm2013/> (26th Mar 2015).
- [5] D.J. Dunstan, A.J. Bushby, Grain size dependence of the strength of metals: The Hall–Petch effect does not scale as the inverse square root of grain size, *Int. J. Plasticity* 53 (2014) 56–65.
- [6] Y. Li, A.J. Bushby, D.J. Dunstan, The Hall-Petch effect as a manifestation of the general size effect, *Proc. Roy. Soc. A* 472 (2016) 20150890.
- [7] J.C.M. Li, Y.T. Chou, Role of dislocations in flow stress grain size relationships, *Metall. Trans. 1* (1970) 1145
- [8] Z.C. Cordero, B.E. Knight and C.A. Schuh. Six decades of the Hall-Petch effect – a survey of grain-size strengthening studies on pure metals, *Int. Mater. Rev.* 61 (2016) 495 – 512.
- [9] V. Bata and E.V. Pereloma, An alternative physical explanation of the Hall–Petch relation, *Acta Mat.* 52 (2004) 657–665.
- [10] J. Aldazabal, J. Gil Sevillano, Hall–Petch behaviour induced by plastic strain gradients, *Mater. Sci. Eng. A365* (2004) 186–190.
- [11] S. Deville, S. Meille, J. Seuba, A meta-analysis of the mechanical properties of ice-templated ceramics and metals, *Sci. Tech. Adv. Mater.* 16 (2015) 043501.
- [12] A.W.F. Edwards, *Likelihood*, John Hopkins University Press 1992.
- [13] Bevan, A.J., *Statistical data analysis in the physical sciences*, Cambridge University Press 2013.
- [14] H. Akaike, New look at statistical-model identification, *IEEE Trans. Automatic Control* 19 (1974) 716–723.
- [15] J.D. Eshelby, F.C. Frank, F.R.N. Nabarro, The equilibrium of linear arrays of dislocations, *Phil Mag.* 42 (1951) 351–364.
- [16] S.D. Antolovich, R.W. Armstrong, Plastic strain localization in metals: Origins and consequences, *Prog. Mater. Sci.* 59 (2014) 1–160.
- [17] E. Walker, A.V. Hernandez, M.W. Kattan, Meta-analysis: Its strengths and limitations, *Cleveland Clinic. J. Medicine* 75 (2008) 431-439.

- [18] H. Cooper, L.V. Hedges, J.C. Valentine (Eds.) *The Handbook of Research Synthesis and Meta-Analysis*, 2nd ed., Russell Sage Foundation, New York, 2009.
- [19] R. Rosenthal, The file drawer problem and tolerance for null results, *Psychological Bulletin* 86 (1979), 638–641.
- [20] J.M.C. Li, Petch relation and grain boundary sources, *Trans. TMS* 227 (1963) 239–247.
- [21] H. Conrad, Effect of grain size on the lower yield and flow stress on iron and steel, *Acta Met.* 11 (1963) 75–77.
- [22] M.F. Ashby, The deformation of plastically non-homogeneous materials, *Phil. Mag.* 21 (1970) 399–424.
- [23] S.B. Nissen, T. Magidson, K. Gross, C.T. Bergstrom, Publication bias and the canonization of false facts, *eLife* 5 (2016) e21451.
- [24] T. Narutani, J. Takamura, Grain-size strengthening in terms of dislocation density measured by resistivity, *Acta Met. Mat.* 39 (1991) 2037–2049.
- [25] A.S. Argon, *Strengthening mechanisms in crystal plasticity*, ch.8.8, Oxford University Press 2008.
- [26] L.M. Brown, Indentation size effect and the Hall-Petch 'Law', *Mater. Sci. Forum*, 662 (2011) 13–26.
- [27] D.J. Dunstan, The size effect in the mechanical strength of semiconductors and metals: Strain relaxation by dislocation-mediated plastic deformation, *J. Mater. Res.* 32 (2017), 4041–4053.
- [28] D.J. Dunstan, 2012, [Critical thickness theory applied to micromechanical testing](#), *Adv. Eng. Mater.* 14 (2012) 942–947.
- [29] E. Wit, E. van den Heuvel, J.–W. Romeyn, "‘All models are wrong...’: an introduction to model uncertainty". *Stat. Neerlandica* 66 (2012) 217–236.

Appendix. Pitfalls of meta-analysis and of standard analysis

The purpose of meta-analysis is to take multiple studies and by combining their results to obtain a greater statistical significance for a result, or, as here, to obtain a result that the original studies did not consider. A review by Walker *et al.* [17] identifies four critical issues for meta-analysis. They are discussed in more detail by Cooper *et al.* [18]. Counterparts arise in standard analysis, such as the effect of canonization discussed by Nissen *et al.* [23].

The file-drawer problem, or publication bias [19] is liable to occur when the meta-analysis is conducted to test the same hypothesis that the original authors were studying. Only, or predominantly, studies with positive outcomes are published, while all the studies with null outcomes languish unpublished in the filing-cabinet. Then, all the meta-study achieves is to confirm the original prejudice according to which positive results were interesting and null results not. That is not a risk here. The original authors did not select for publication only those datasets which fitted well with Eq.2. They might have rejected those that did not fit Eq.1, but that would not matter to us. What matters is that they did not – could not – select data for publication according the fit with the hypothesis that we are testing.

The comprehensiveness of the search for studies also matters for if it is not comprehensive there is scope for selection bias in the studies selected for inclusion in the meta-study. Walker *et al.* [17] do not emphasise, though, that if the search not comprehensive, but is random with respect to the hypothesis under test, having fewer studies merely lowers the statistical significance of the result of the meta-analysis but does not invalidate it in any other way. That is clearly the case here, for the same reason as this study does not risk publication bias.

Canonization, or the inverse file-drawer problem, in standard analysis, occurs when a fact becomes taken for granted rather than treated as an open hypothesis [23]. It may be false but consistent with the reported data as the inverse square-root dependence of Eq.1. Yet subsequent reviewers may make it hard to publish interpretations that call it into question, or subsequent authors may prefer to use it than to challenge it for fear of the negative reaction of reviewers. Papers, too, can become canonized – this is related to the “often cited, rarely read” phenomenon, in which repeated citation without critical reading of the original leads to the paper being considered to demonstrate something which in fact it does not. A relevant case here is the highly-cited paper of Narutani and Takamura [24], a very nice study of the Hall-Petch effect, reporting resistivity measurements to access dislocation densities, as well as the usual stress-strain data. The paper supports the standard Hall-Petch expression, with the refinement of an increase in the exponent x from $\frac{1}{2}$ at low plastic strains to 1 at high plastic strains in accordance with the Ashby equation; see e.g. Argon [25] and Brown [26].

Actually, measurements such as those reported by Narutani and Takamura [24] are not able to determine the exponent x , let alone a change in it (see Baldwin [3]). Their Fig.12 is a Hall-Petch plot of their stress data, with linear guides to the eye drawn through the 0.2%, 2.5% and 5% strain data, and concave guides through the 10%, 15% and 20% data. The latter are quite convincing, but highly exaggerated, being approximately sixth-degree polynomials rather than the parabolae expected for a d^{-1} dependence on a $d^{-\frac{1}{2}}$ abscissa. In Fig.A1 we plot the data from their Fig.12, together with the three fits, using $x = \frac{1}{2}$, 1 and with x as a free fitting parameter. The fitted values are given next to each dataset. It is clear that these data cannot distinguish the three fits. Nor does the resistivity data, converted to stress using the information, the linear fits, in their Fig.5 and Fig.8 and shown in the Hall-Petch plot of Fig.A1, add anything. They are in quite remarkable agreement with the stress-strain data at the smaller grain sizes, and consistent with the standard Hall-Petch exponent within about $1\frac{1}{2}\sigma$ in the fitted exponent x . Given especially our finding above that plastic strain is not a significant factor, this shows well the pitfall in standard analysis of undue dependence upon a single paper compared with the overview that meta-analysis provides.

[Fig. A1 near here]

Heterogeneity of results, or not comparing like with like, risks burying a few positive results from well-focused studies under scattered results from many less relevant studies. That is not an issue here. All the datasets that we used (with one exception [24]) reported good fits to Eq.1.

Availability of relevant information is the third key issue, and it does apply here. Among these datasets are of course wide variations of techniques such as grain-size

measurement and characterisation of texture. Such variations were not generally fully reported by the original authors and the data were not then and cannot now be corrected in any way for them. However, such variations did not affect the validity of the datasets as published or later used as support for Eq.1, and no more do they affect their validity under meta-analysis as evidence of a behaviour which refutes Eq.1. Indeed, selection or correction of the raw data is dangerous in meta-analysis because of the risk of introducing bias. Some authors gave information about, for example, measurement of grain size, while others did not. This is not a risk factor providing that (1) there is little likelihood that the reported information or lack of it is correlated any of the factors, and (2) that no attempt is made to correct some data in the light of this information while other data cannot be corrected for lack of the information.

Analysis of data is a rather technical issue that does not concern us here, for it covers issues such as data-mining, in which a large body of data, tested for a very large number of correlations, will by chance give some false-positive outcomes among the very large number of true-negatives. Here we are looking for and finding specific outcomes predicted by theory.

Good physics: The previous point raises a final criticism of meta-analysis, made by a referee of a previous version of this paper: good physics does not arise out of statistical analyses of large datasets. Here, of course, the physics does not arise out of the statistical analysis. The physics arises out of the Orowan-Matthews ideas about dislocation curvature, or out of the Eshelby-Frank-Nabarro pile-up ideas [27]. Statistics are used merely to choose between them, in as rigorous a way as possible. It is for this reason that we do not use a coefficient of determination such as R^2 to test Eq.1 or Eq.3, but consider, simply, the question, what model is most probable, given the data?

Table 1. Results of fitting randomized datasets with various Maximum Likelihood (ML) probability distributions, Gaussian (G), equivalent to a least-squares (LS) fit, Lorentzian (Lor), double-Gaussian (DG) and Gaussian plus flat (GF).

	ML/MS G	ML Lor	ML DG	ML GF
$P(r_i)$	$\frac{1}{\sqrt{2\pi}\sigma^2} e^{-r_i^2/2\sigma^2}$	$\frac{1}{\pi} \frac{\gamma}{r_i^2 + \gamma^2}$	$fG(\sigma_1) + (1-f)G(\sigma_2)$	$fg + (1-f)G(\sigma)$
500 LS dummies	$a: -0.003 \pm 0.092$			
Dummy D1	$\ln L: 10.0$ $f: 0$ $a: -0.056 \pm 0.096$ $\sigma: 0.29$	$\ln L: 56.6$ $f: 0$ $a: -0.018 \pm 0.024$ $\gamma: 0.034$	$\ln L: 61.2$ $f: 0.14 \pm 0.5$ $a: -0.010 \pm 0.025$ $\sigma_1: 0.07; \sigma_2: 0.8$	$\ln L: 68.5; g: 1$ $f: 0.15 \pm 0.05$ $a: -0.011 \pm 0.025$ $\sigma: 0.062$
$\log_{10} D1$	$\ln L: -38.0$ $f: 0$ $a: -0.17 \pm 0.21$ $\sigma: 0.65$	$\ln L: -29.2$ $f: 0$ $a: -0.072 \pm 0.11$ $\gamma: 0.16$	$\ln L: -26.5$ $f: 0.42 \pm 0.12$ $a: -0.076 \pm 0.11$ $\sigma_1: 0.21; \sigma_2: 1.0$	$\ln L: -25.1$ $f: 0.09 \pm 0.03; g: 3$ $a: -0.078 \pm 0.12$ $\sigma: 0.27$

Table 2. Fitting parameters for the various factors of Figs 2-4 using the Maximum Likelihood method with the Gaussian plus Flat pdf. The fraction of outliers is f and the maximum log-likelihood is $\ln L$.

a. Fits to the Hall-Petch inverse-square root equation				
Abcissa	Slopes a	Slopes α	Outliers f	$\ln L$
Year, Y	$a_Y = 0.032 \pm 0.050$	$\alpha_Y = 3.3 \times 10^{-4} \text{ yr}^{-1}$	0.14 ± 0.05	69
Strain, ε_{pl}	$a_\varepsilon = 0.038 \pm 0.026$	$\alpha_\varepsilon = 0.126$	0.15 ± 0.06	70
Schmid factor anisotropy, Δm	$a_m = 0.039 \pm 0.038$	$\alpha_m = 0.123$	0.14 ± 0.06	70
Purity, p	$a_p = 0.067 \pm 0.015$	$\alpha_p = 0.033$	0.16 ± 0.06	76
Bulk strength, σ_0	$a_\sigma = 0.059 \pm 0.023$	$\alpha_\sigma = 20.3$	0.15 ± 0.06	72
Anisotropy, C	$a_C = 0.086 \pm 0.020$	$\alpha_C = 0.0117$	0.19 ± 0.06	75
Stacking fault energy, γ	$a_\gamma = 0.088 \pm 0.015$	$\alpha_\gamma = 5.3 \times 10^{-4} \text{ m}^2 \text{ mJ}^{-1}$	0.14 ± 0.05	84
Grain size, d_{ISR}	$a_d = 0.101 \pm 0.029$	$\alpha_d = 3.9$	0.15 ± 0.05	74
b. Fits to the critical thickness $\log d$ over d equation				
Abcissa	Slopes a	Slopes α	Outliers f	$\ln L$
Year, Y	$a_Y = 0.85 \pm 0.55$	$\alpha_Y = 8.6 \times 10^{-3} \text{ yr}^{-1}$	0.15 ± 0.05	-89
Strain, ε_{pl}	$a_\varepsilon = 0.40 \pm 0.28$	$\alpha_\varepsilon = 1.34$	0.15 ± 0.05	-89
Schmid factor anisotropy, ΔM	$a_M = 0.97 \pm 0.40$	$\alpha_M = 3.1$	0.15 ± 0.05	-87
Purity, p	$a_p = 0.71 \pm 0.17$	$\alpha_p = 0.36$	0.15 ± 0.05	-82
Bulk strength, σ_0	$a_\sigma = 0.45 \pm 0.33$	$\alpha_\sigma = 110$	0.15 ± 0.05	-89
Anisotropy, C	$a_C = 0.81 \pm 0.26$	$\alpha_C = 0.107$	0.16 ± 0.05	-86
Stacking fault energy, γ	$a_\gamma = 0.81 \pm 0.18$	$\alpha_\gamma = 4.8 \times 10^{-3} \text{ m}^2 \text{ mJ}^{-1}$	0.15 ± 0.05	-81
Grain size, d_{ISR}	$a_d = 0.81 \pm 0.33$	$\alpha_d = 33$	0.15 ± 0.05	-87

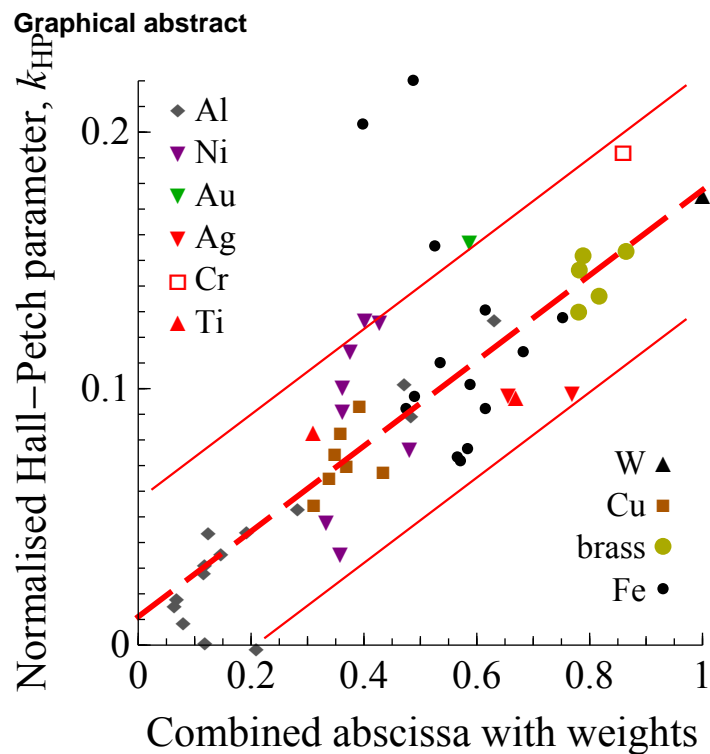
Table 3. Fitting parameters for weighted combinations of factors. Fixed fitting parameters are in bold. In the first column, equal weights are given to all factors. In the second column, weights are optimized for the highest $\ln L$. Then each factor in turn has been eliminated by giving a fixed weight of zero and $\ln L$ has been optimized with respect to the other weights. The final rows gives the Akaike Information Criterion (AIC) [15] and the Bayesian Information Criterion (BIC) [29] values for each model.

a. Fits to the Hall-Petch inverse-square root equation										
Factors		w_i	w_i	w_i	w_i	w_i	w_i	w_i	w_i	w_i
Year	1	-0.5 ± 1.4	0	-0.5	-1.0	-0.6	0.5	-0.7	-0.9	0.8
Strain, ε_{pl}	1	2.6 ± 1.5	2.4	0	2.9	1.3	2.5	1.1	4.4	2.9
Purity, p	1	0.6 ± 0.4	0.6	0.7	0	0.6	0.9	0.7	1.0	1.2
Bulk strength, σ_0	1	-1.0 ± 1.0	-0.9	0.7	-1.2	0	-0.9	0.6	-2.1	0.4
Anisotropy, C	1	1.1 ± 0.4	1.1	1.1	1.3	1.1	0	1.4	1.2	1.9
Stacking-fault energy, γ	1	0.9 ± 0.5	0.8	0.7	1.1	0.7	1.7	0	1.5	2.0
Schmid factor anisotropy, Δm	1	2.1 ± 1.1	2.0	2.9	2.8	2.6	0.8	3.0	0	-1.1
Grain size, d_{ISR}	1	2.2 ± 0.5	2.1	2.4	2.1	2.2	2.5	2.0	2.8	0
“Outliers”, f	0.19	0.15	0.15	0.15	0.15	0.15	0.15	0.18	0.15	0.28
Slope, a	0.142	0.156 ± 0.06	0.157	0.155	0.159	0.156	0.162	0.150	0.148	0.146
Gaussian width, σ	0.024	0.023	0.023	0.024	0.024	0.023	0.025	0.022	0.024	0.017
Log-Likelihood, $\ln L$	92	101.4	101.4	99.8	99.7	100.9	97.5	98.4	100.1	89.3
AIC = $2n - 2 \ln L$	-175	-181	-183	-180	-179	-182	-175	-177	-180	-159
BIC = $n \ln 61 - 2 \ln L$	-167	-158	-162	-159	-158	-161	-154	-156	-159	-138

Table 3b. Fitting parameters as Table 3a but for $d^{-1}\ln d$

Factors	w_i	w_i	w_i	w_i	w_i	w_i	w_i	w_i	w_i	w_i
Year, y_r	1	1.0 ± 1.0	0	0.9	0.9	1.0	1.2	0.8	1.0	1.8
Strain, ε_{pl}	1	2.0 ± 2.5	2.1	0	2.3	1.8	1.6	0.7	2.9	5.1
Purity, p	1	0.8 ± 0.6	0.8	0.8	0	0.8	0.9	0.9	1.2	1.0
Bulk strength, σ_0	1	-0.2 ± 3.0	0.1	2.0	0.8	0	0.3	1.8	-1.0	-4.3
Anisotropy, C	1	0.8 ± 0.8	1.0	0.6	1.0	0.8	0	1.1	0.4	1.3
Stacking-fault energy, γ	1	1.0 ± 0.7	1.1	0.8	1.4	1.0	1.4	0	1.5	1.5
Schmid factor anisotropy, Δm	1	1.2 ± 0.8	1.2	1.2	1.6	1.2	0.9	1.3	0	1.7
Grain size, d_{ISR}	1	1.5 ± 0.8	1.8	1.7	1.6	1.5	1.8	1.4	2.0	0
“Outliers”, f	0.15	0.15 ± 0.05	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.15
Slope, a	1.76	1.77 ± 0.13	1.56	1.95	1.75	1.78	1.74	1.99	1.68	1.53
Gaussian width, σ	0.39	0.38 ± 0.04	0.38	0.38	0.39	0.38	0.38	0.39	0.39	0.38
Log-Likelihood, $\ln L$	-72.3	-71.2	-71.7	-71.6	-72.4	-71.2	-71.8	-72.5	-72.3	-72.8
AIC = $2n - 2\ln L$	153	164	163	163	165	162	164	165	165	165
BIC = $4.1n - 2\ln L$	161	188	184	184	186	183	185	186	186	187

Graphical abstract



$$\sigma = \sigma_0 + \frac{k_{HP}}{\sqrt{d}}$$

The abscissa contains weighted factors:

w_1 year of publication

w_2 plastic strain

w_3 plastic anisotropy

w_4 composition (purity)

w_5 bulk strength

w_6 elastic anisotropy

w_7 stacking fault energy

w_8 grain size

Only *grain size* is statistically significant
in determining k_{HP} values

Fig1 with caption

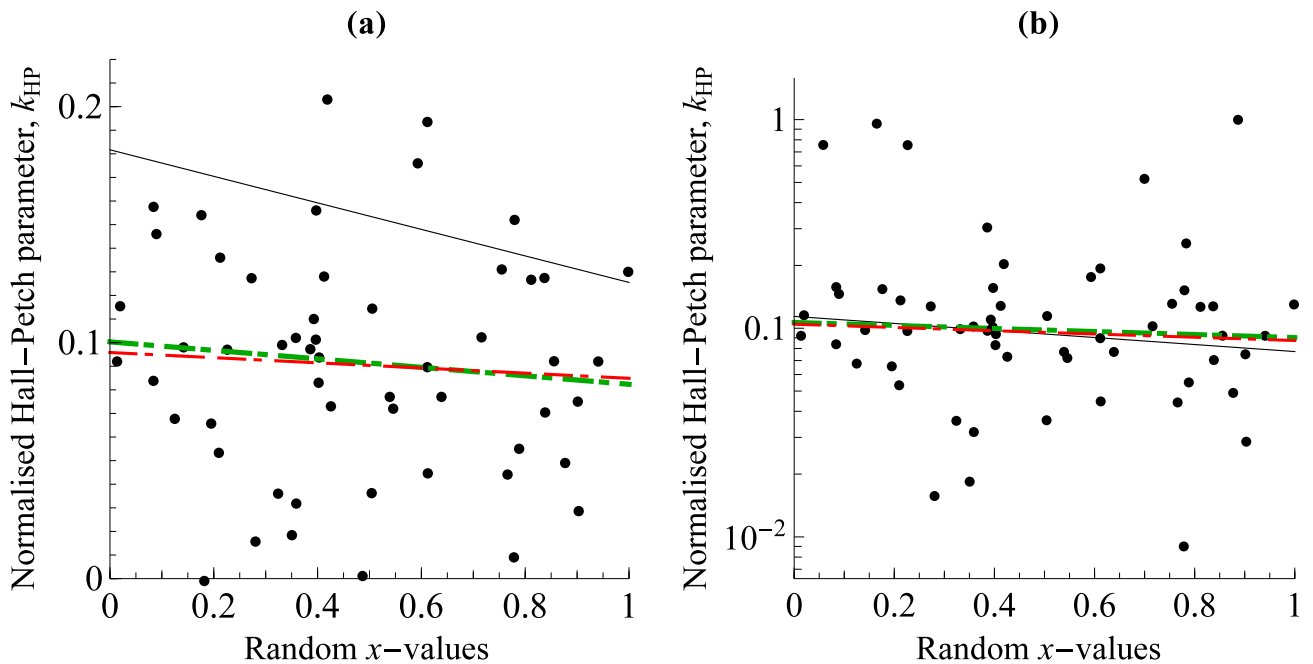


Fig.1. (Colour on-line.) The 61 values of the Hall-Petch parameter k_{HP} on (a) a linear plot and on (b) a logarithmic plot against a random abscissa in the range 0-1. The extreme high values are not shown in (a) but may be seen in (b); similarly the extreme low and negative values are not shown in (b) but may be seen in (a). Fits are shown by a classic least-squares (LS) procedure (thin solid lines) and by Maximum Likelihood methods using the Lorentzian pdf (thick chain-dotted green lines) and the Gaussian plus flat pdf (GF, thick dashed red lines) with the values given in Table 1.

Fig2 with caption

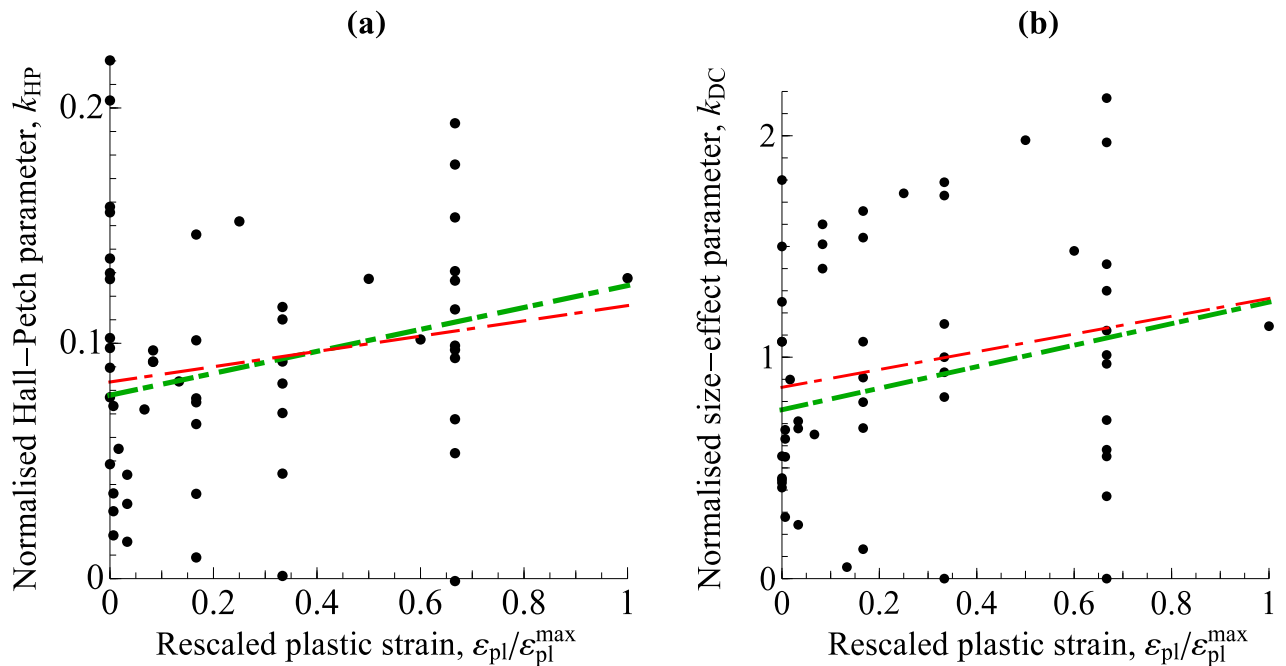


Fig.2. (Colour on-line.) Plots and fits against plastic strain. In (a) the data are the values of k_{HP} and in (b) the data are the values of k_{DC} . The thick chain-dotted green lines are the fits using the Lorentzian pdf and the thick dashed red lines are the fits with the GF pdf.

Fig3 with caption

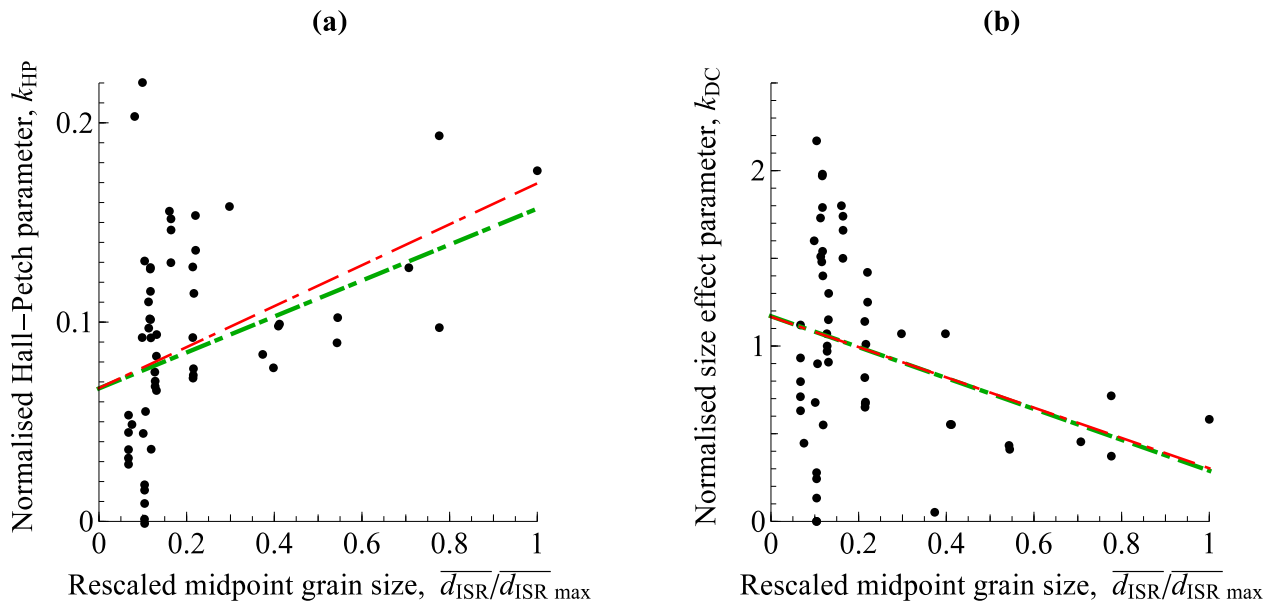


Fig.3. (Colour on-line.) Plots and fits against the midpoint of the inverse square roots of the grain sizes of each dataset. In (a) the data are the values of k_{HP} and in (b) the data are the values of k_{DC} . The thick chain-dotted green lines are the fits using the Lorentzian pdf and the thick dashed red lines are the fits of the GF model.

Fig4 with caption

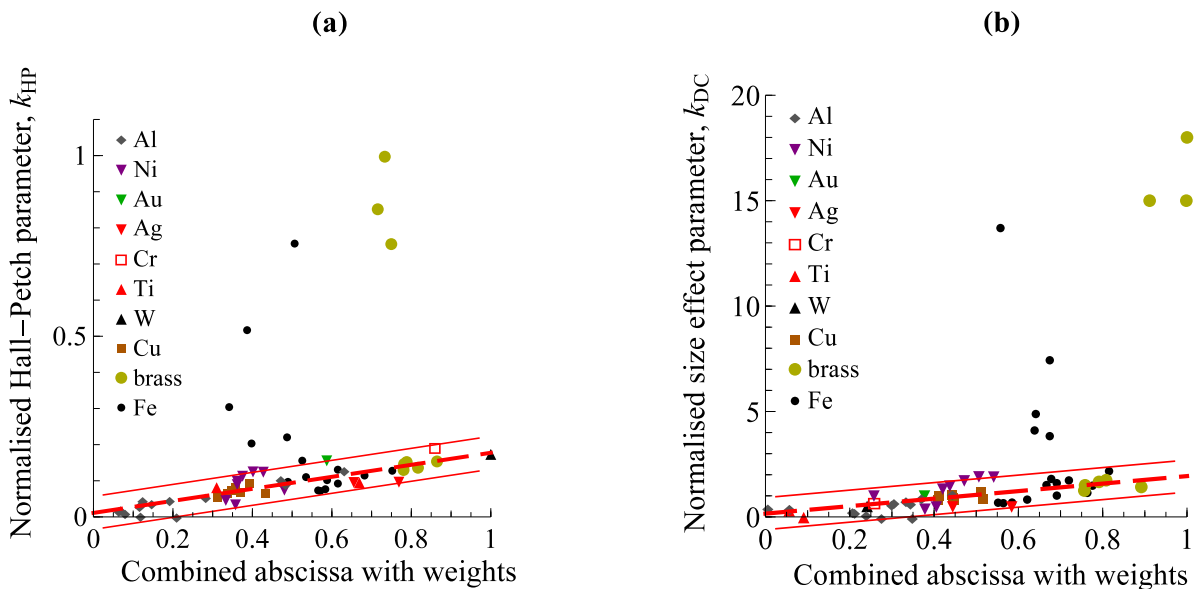


Fig.4. (Colour on-line.) Plots and fits of (a) k_{HP} and (b) k_{DC} against the weighted combination of all eight factors of the third columns of Table 3a and 3b. The thick dashed red lines are the fits of the GF model. The thin red solid lines show the width (two-sigma, within which 95% of the data are expected to lie) of the Gaussian distribution of the mainstream data around the GF fit.

Appendix FigA1 with caption

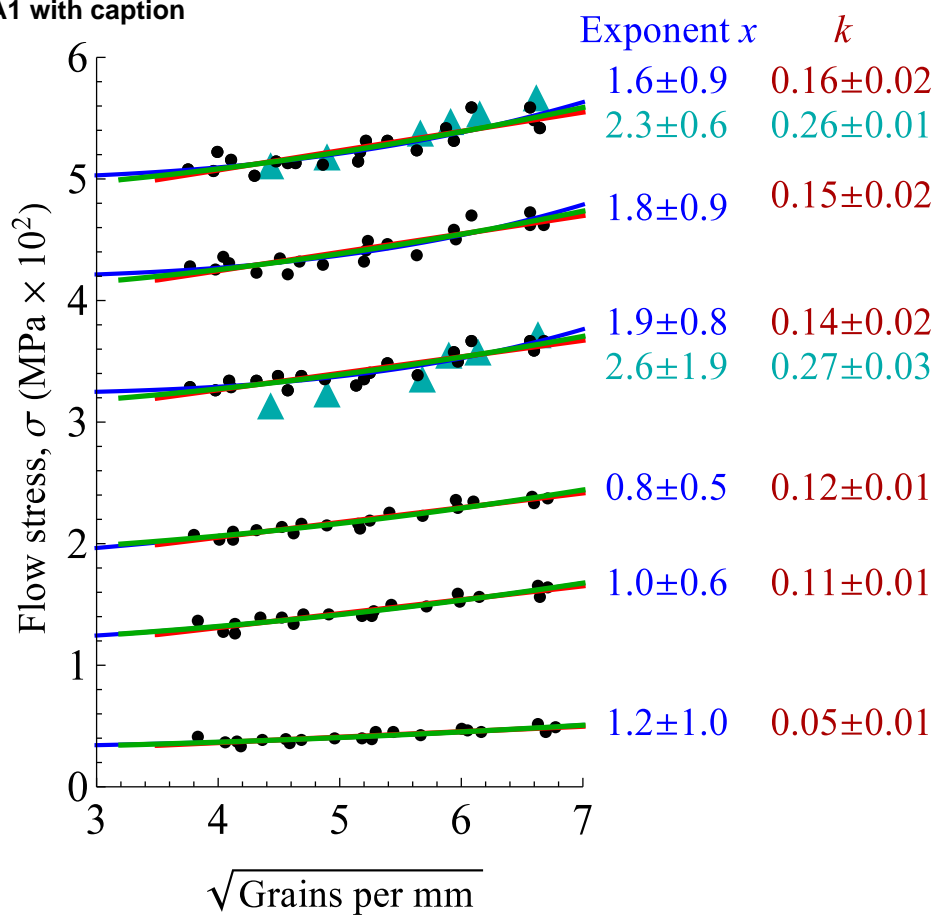


Fig.A1. (Colour on-line.) The stress-strain data of Narutani and Nakamura [24] are plotted (small black points) for nickel under tension at 77K, for strains of 0.02% (bottom), 2.5%, 5%, 10%, 15% and 20% (top). The stress-strain data that can be deduced from their electrical measurements are shown, for 10% and 20% strain (cyan triangles). Fits to the stress data are shown, using Eq.1 with exponent $x = \frac{1}{2}$ (red), with $x = 1$ (green), and with x as a free fitting parameter (blue). The values of the fitted exponents x and of the Hall-Petch parameter k (for $x = \frac{1}{2}$) are given; x is dimensionless while k is in units of $\text{MPa} \cdot 10^2 \text{ mm}^{\pm 1/2}$. The values for the electrical data are in cyan.